MEMS- and Probe-Based Mass Data Storage in Conventional CMOS

Michael S.-C. Lu*, Seungook Min[b], Shien-Der Tzeng[a], and Shangjr Gwo[a],
*Department of Electrical Engineering and Institute of Electronics Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, China
[a] Department of Physics, National Tsing Hua University, Hsinchu 30013, Taiwan, China
[b] Department of Material Science, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
*Email: sclu@ee.nthu.edu.tw

Abstract

Non-volatile mass data storage using MEMS-based actuation and probe-based read/write has great potential to replace conventional semiconductor memories because of the ever achieved high areal density. From the standpoint of system implementation, the complexity for integration of all functional electronics in such a micro disk drive is enormously reduced by using the CMOS-MEMS as the core implementing technology. In this paper, we will present the latest progresses on the MRM-based and the EFM-based storage devices that are currently under development.

1. Introduction

The development of a probe-base mass data storage device requires multi-disciplinary research efforts from the fields of MEMS and scanning probe microscopy. The former is known for its capability to fabricate micrometer-sized electromechanical devices that possess higher mechanical bandwidth and stiffness than their macro-counterparts. Those properties, from the control systems perspective, are translated to faster response time, finer positioning accuracy, and better immunity to external disturbances (e.g. vibrations and shocks). The latter, such as STM [1], AFM [2], SThM [3, 4], MFM [5,6], etc, provides the capability to render nanometer-scale solution of detected physical properties (e.g. topography in STM and AFM, temperature distribution in SThM). Some scanning mechanisms can be utilized to form stored bits for non-volatile mass data storage.

MEMS-based mass data storage featuring low power, high storage capacity, and miniaturized size can be used in mobile computing devices to provide a quantum leap for the storage capacity of non-volatile memory which is now mainly served by flash-memory chips. Based on the progress of lithographical process predicted by the ITRS (λ = 32 nm in 2013), it is not likely to get several gigabytes of storage capacity on a cm^2 size flash chip until ten years later (each bit cell = 3 to 4λ^2). Instead, probe-based data storage can be one of the best solutions with an array of micro tip-cantilevers being simultaneously actuated for reading and writing of tiny bits at high throughput.

One of the best-developed probe-based data storage devices is the IBM Millipede [7], which has demonstrated 1 Tb/in^2 storage density using a passive 32 x 32 tip-cantilever array. The writing mechanism in this system uses probe tips to thermally melt pits in a polymer media. The reading of 0 or 1 depends on the change of thermal conductance when tips are either confined in pits or on a flat surface. For data access, all the tip-cantilevers are moved up to the media by a die-level actuation. Each cantilever is passive with no individual actuation and position servo. The reliability and wear of tips may be an issue for long-term operation.

In this paper, we will report the recent developments of two mass data storage devices that are based on the MRM (Magneto-resistive Microscopy) and the EFM (Electrostatic Force Microscopy) scanning mechanisms. The significant difference from the IBM Millipede approach is that the scanning operates in non-contact mode, which requires tip-level position servo but ensures a better tip reliability. In the following sections we will discuss the aspects of reading/writing mechanisms, fabrication, actuator design, and position servo regarding these two approaches.

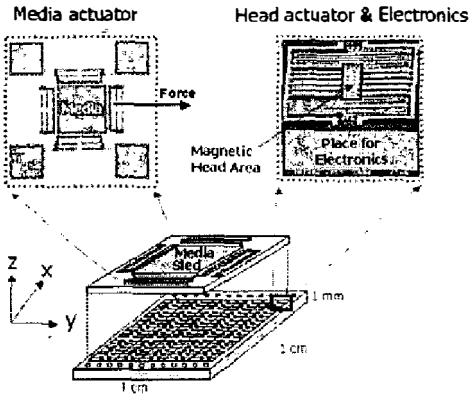

Figure 1 System schematic of the MEMS- and probe-based micro disk drive.

0-7803-8511-X/04/$20.00 ©2004 IEEE.

2. MRM-Base Mass Data Storage

2.1 System Architecture

The system-level description of the MRM-based micro disk drive can be found in [8]. As shown in Figure 1, the device consists of two bonded dies: the upper die has a x-y actuated media actuator with deposited magnetic media, and the bottom CMOS-MEMS die has an 80 x 80 array of z-actuated read/write head actuators and their associated circuits for read/write channels and position servo.

2.2 Reading and Writing Mechanisms

Various reading and writing processes have been considered; for example, in the MFM (Magnetic Force Microscopy)-like approach, writing is achieved with a hard magnetic tip approaching the medium for magnetization. The written bit is read by approaching the tip and sensing the magnetic force between the tip and the bit using a cantilever. In this case the distance for reading must be larger than that for writing in order not to change the magnetization of the bit.

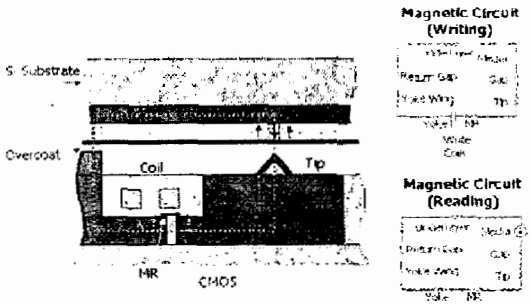

Figure 2 Cross-sectional view of the read/write probe head.

A soft magnetic tip is used for reading and writing in the MRM-based micro disk drive. Writing is achieved by generating a current induced field, and reading is realized by using an anisotropic magnetoresistive (AMR) sensor, basically a single layer of permalloy in our design. The permalloy AMR sensor can be easily saturated with very low magnetic field (2 Oe or less) and shows about 2% resistive change. Compared with the MFM approach, the read/write speed for MRM is orders of magnitude faster because it is not limited by the mechanical bandwidth of the tip actuator. The cross-section view in Figure 2 illustrates the magnetic yoke-type head design on top of CMOS, and the associated magnetic reading/writing circuits for the proposed perpendicular magnetic recording. For reading, the magnetic flux from the medium is guided through the magnetic tip, yoke and the magnetoresistive (MR) sensor. The flux divides when reaching the yoke gap, so one part passes through the MR sensor and the other through the yoke gap. The resistance of the MR sensor changes with the applied magnetic flux as the read signal. The divided flux through the MR sensor and the yoke gap, after recombination, is guided to the soft under layer in the medium to close the path. The yoke wings are an attempt to increase the efficiency of the magnetic circuit, bringing the permeable material closer to the medium for a better return path.

The estimated writing field is around 400 kA/m. In the case of perfect head efficiency, a minimum current of 3 mA is required, if three turns of coils are used. Standard CMOS Al interconnects can handle about 1 $mA/\mu m^2$. Thus the actuator design needs to use 3-μm wide spring to handle the current, which, on the other hand, increases the actuation voltage. To lower the write current, local heating of the medium has been studied [9] by applying a tunneling current through the tip to reduce the coercivity of the medium.

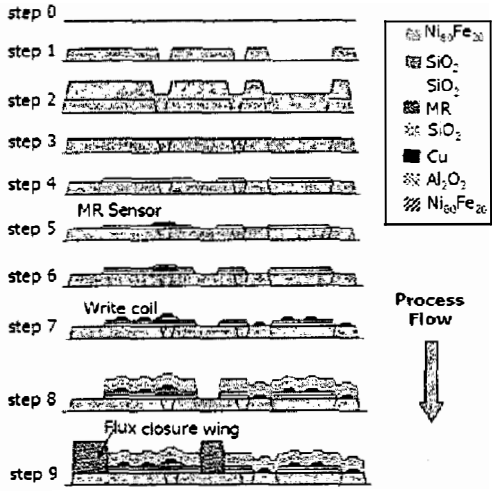

Figure 3 Cross-sectional view of probe head fabrication.

The critical factors that determine the MR sensitivity are the yoke gap under the MR sensor and the sensor alignment. A silicon dioxide layer is deposited between the MR sensor and the permalloy yoke for electrical insulation. The layer thickness should be tightly controlled with respect to the yoke gap width, in order to let the magnetic flux pass through the yoke gap and the MR sensor simultaneously for optimal operation. With a thicker insulation layer than the yoke gap, little magnetic flux from the medium will pass through the read sensor. Conversely, a very large gap in the yoke produces a large value of reluctance and degrades the head efficiency. Therefore the gap should be the lithographic minimum dimension of the head process

and the insulation layer should be thinner than the gap width.

For process integration with the CMOS, a series of low-temperature processing steps for head fabrication has been devised, as depicted in Figure 3. On the top of the CMOS (step 0), a 2-μm thick permalloy for the yoke layer is deposited by sputtering, followed by patterning (step 1). The patterned permalloy is covered with a 3-μm to 5-μm silicon dioxide in step 2, and then planarized with the chemical mechanical polishing (CMP) in step 3 for the subsequent MR deposition. The planarized silicon dioxide surface is then covered with a 0.1-μm silicon dioxide that is patterned by the reactive ion etching (RIE) in step 4. MR layer is sputtered and patterned in step 5. After another thin silicon dioxide deposition and patterning in step 6, the 0.5-μm thick copper coils are electroplated in step 7. The copper wire is covered with another silicon dioxide layer in step 8. Finally a 3–μm thick permalloy layer for the flux closure wing is sputtered and patterned (step 9). The tip is deposited in a patterned recess as in the Spindt-type tip fabrication. For release of the magnetic probe head along with the actuator, a thick photoresist is spun on top and used as the mask for the oxide and the silicon etch [10], followed by removal of photoresist using oxygen plasma.

The low-temperature MR thin film [11] was deposited and patterned on a structure that was intended for test of process integration. Figure 4(a) shows the increase of MR resistance due to dimension loss after the oxide etch and the silicon etch. An approximate 0.3% MR change was measured after the silicon etch as shown in Figure 4(b). The relatively low resistive change compared with the normal 2% is due to the series-connected large lead resistance through the long narrow spring. The large saturation field is due to the incompatibilities between the permalloy yoke design of the test structure and the MR probe station. The original read-write magnetic head was designed to allow most of the magnetic flux to pass through the MR sensor, but in this MR test structure, most of the magnetic flux was shunt to the yoke layer so the saturation field becomes much higher.

The low signal to noise ratio (SNR) in the AMR head is mainly due to the high routing resistance through the meandering spring. The routing resistance can be reduced with less turns of springs being used, provided that a high-voltage CMOS process is available for actuation. The tunneling magnetoresistive (TMR) sensor is also currently under investigation in order to gain a high SNR. The TMR sensor is operated based on the spin-dependent tunneling effect that uses tunneling current through an insulation layer during reading. It provides a much higher nominal resistance than that of the AMR sensor. Also, the TMR sensor shows a high MR ratio (between 15 % and 50 %) even in a small magnetic field.

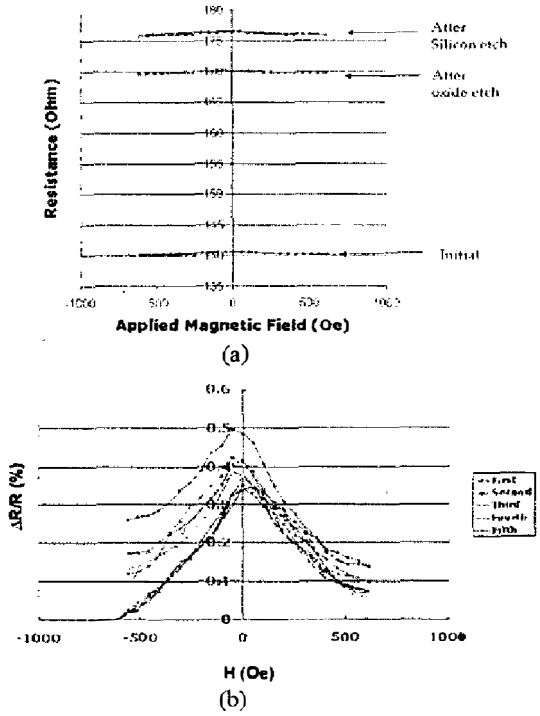

Figure 4 MR measurement after silicon etch.

2.3 Position Servo of Tip Actuators

Each tip actuator is capable of performing position sensing and servoing to keep the tip about 10 nm away from the media for data access. A total of 7 interconnects is required for signal routing in the actuator, including capacitive sensing, electrostatic actuation, reading (×2), writing, thermally-assisted writing, and ground. The needs for routing and circuit integration make the CMOS-MEMS technology an ideal choice for realizing the tip/cantilever array. The tip actuator as shown in Figure 5(a) has two anchored springs, two capacitive-sensing plates, and one actuation plate. Four leads are embedded in each spring, thus a total of 8 wires can be provided.

To reduce power consumption for actuation of the 80 × 80 cantilevers, parallel-plate electrostatic actuation is used in favor of other mechanisms. For data access, however, we have to face the electrostatic pull-in limit as in conventional parallel-plate actuators. The charge-control approach by using a switched-capacitor circuit [12] has demonstrated a stable travel distance more than 80% of the gap. The actuation capacitance in our design is nonetheless smaller than 10 fF, which makes the switched-capacitor approach an inappropriate choice. We have, on the other hand, demonstrated the use

of voltage feedback [13] to extend the travel range up to 60% of the gap, nearly twice of the pull-in limit. The control system around the tip cantilever, as shown schematically in Figure 5(b), consists of a linear stabilizing controller that has the minimum complexity for duplication in an array. The position of the actuated plate (C_{act}) is detected by using the sensing capacitances C_{s1} and C_{s2} in a continuous-time voltage-sensing circuit. Assuming an initial gap of 3 μm and a tip height of 0.5 μm, a tip displacement up to 80% of the gap is desired to access the stored bits. The failure of a fixed linear controller is due to its insufficiency to adjust in accordance with the changing unstable pole along the motion trajectory. A very viable approach in the future for improvement of the phase margin, especially in the gap-closing region, is to use a linear controller that has a variable gain along the motion trajectory.

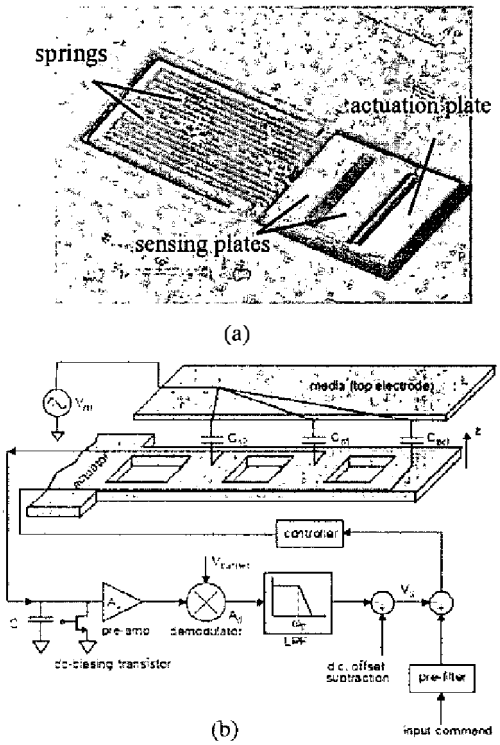

Figure 5 (a) SEM of released tip actuator. (b) Schematic of the control system around a tip actuator.

3. EFM-Based Mass Data Storage

3.1 Introduction

Since its first introduction in 1967 by Wegener *et al.* [14], metal-nitride-oxide-silicon (MNOS) structure and its various derivatives have remained as one of the state-of-the-art techniques for nonvolatile semiconductor memories. Inspired by the concept of movable local metal electrode for controlling and sensing trapped charge to achieve a high storage density [15], R. C. Barrett and C. F. Quate successfully applied scanning capacitance microscopy (SCM), a variant of contact-mode AFM, to manipulate charges in NOS samples [16]. Subsequent works have confirmed similar ideas using different scanning probe concept or dielectric structure [17-24]. In our recent work on NOS structures [25], we have demonstrated the possibilities of ultra high areal density (500 Gbit/in^2), fast writing speed (3.8 mm/sec, limited only by instrument), low writing voltage (5-10 V), and practically unlimited charge retention time at room temperature by using electrostatic force microscopy.

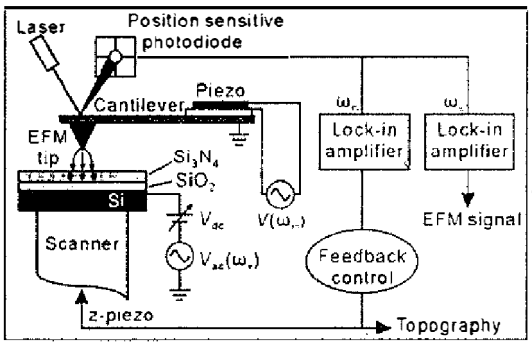

Figure 6 A dual-modulation scheme for EFM signal detection.

The read/write/erase steps in our experiments are achieved by EFM under the high-vacuum conditions (10^{-6}-10^{-7} Torr), which allows higher charge sensitivity and spatial resolution [26]. Also, it prevents the local charging process from the effects of probe-induced anodic oxidation [27] and tip wearing. Both the write and erase steps are easily achieved by injecting electrons or holes from the conductive EFM tip with a positive or negative sample bias. The write/erase steps are mainly based on the Fowler-Nordheim field-emission mechanism. The charge sensing using an AFM probe in close proximity is relatively complex due to the interfering van der Waals force and various chemical interactions. To separate the electric force information from these effects, a dual-modulation scheme as shown in Figure 6 was used. A near-resonant mechanical modulation ($\omega_m \approx \omega_0$) was applied to the cantilever and a non-resonant electric modulation (ω_e) was applied to the sample bias. The sample bias has tunable dc and ac components (V_{dc} and V_{ac}). The electric field generated by the electric modulation is typically quite small and tunable by adjusting V_{dc}, V_{ac}, and ω_e. In the experiments, the output of the cantilever vibration detector (a position-sensitive photodiode) is fed to two lock-in amplifiers. The first harmonic signal of ω_m (obtained

from the internal lock-in) is used for the Z-control feedback of the AFM system. Signals from the Z-control feedback and from the second external lock-in amplifier (first harmonic signal of ω_e) are recorded simultaneously as topography and EFM images.

3.2 CMOS-MEMS Cantilever Sensing and Servo

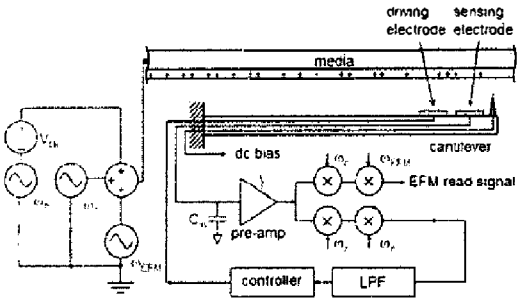

Figure 7 Cantilever-level position sensing, servoing, and EFM read/write.

The schematic shown in Figure 7 illustrates the implementation of position sensing, servoing, and EFM read/write around a CMOS-MEMS cantilever. Similar to the aforementioned procedure, the operation starts with the electrostatic excitation of the cantilever at resonance $\omega = \omega_n$ (around 100 kHz). The resonant amplitude is used as the sensed signal to keep a constant scanning height in a servo loop. EFM reading is achieved with the use of another exciting signal at $\omega = \omega_{EFM}$ ($\omega_{EFM} < \omega_n$). The induced displacement change due the stored charges is much smaller than the cantilever resonant amplitude at $\omega = \omega_n$, thereby eliminating the need for feedback control. The cantilever resonant amplitude at $\omega = \omega_n$ and ω_{EFM} is sensed capacitively using a single-ended scheme, in which the sensing capacitance between the sensing electrode and the top media is less than 1 fF. Consequently the pre-amp input capacitance has to be minimized in order to avoid significant signal attenuation. Due to the extremely small sensing capacitance, any other amplifier topology that employs the concept of capacitive feedback is inappropriate because the value of parasitic capacitances can overwhelm the sensing capacitance. To avoid the sensed signal being overwhelmed by feedthroughs from the actuation electrode to the sensing electrode, a modulation signal at $\omega = \omega_c$ is applied on the media to modulate the respective cantilever motions to frequencies at $\omega = \omega_c \pm \omega_n$ and $\omega_c \pm \omega_{EFM}$, which subsequently need two demodulations to obtain base-band signals.

The CMOS-MEMS fabrication [10] allows many CMOS interconnect layers to be used for routing of the required signals as shown in Figure 7. The pre-amplifier can be placed closed to the cantilever (~ 20 μm) to reduce the parasitic capacitance seen at the sensing node. The most dominating parasitic capacitance would result from the routing within the long cantilever structure. By inserting a metal shield adjacent to the signal line which is driven by the pre-amplifier, the total input capacitance expects to be minimized to below 0.5 fF. Figure 8 illustrates the schematic of the buffer amplifier design. The high-impedance current source consolidates the low impedance looking into the source of input transistors, resulting in a low C_{gs}. C_{gd} is reduced by use of the high-impedance current mirror as the load. The designed spring constant of the cantilever is around 1 N/m, allowing a 1-nm motion from the sensed charges. The control problem, compared with that in the MRM-based storage device, is relatively simpler because the feedback loop mainly performs set-point regulation to maintain constant resonating amplitude for the cantilever.

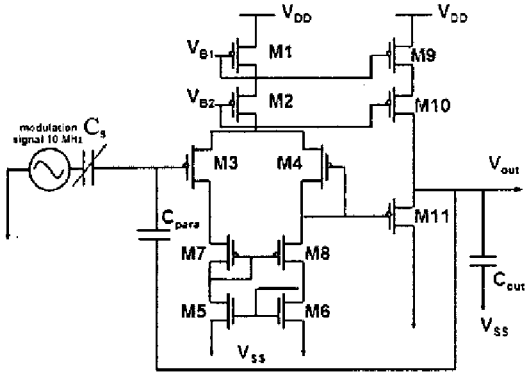

Figure 8 Schematic of the ultra-low input capacitance pre-amplifier

3.3 Experiment

The commercial AFM/EFM probes used in this work have a typical tip radius less than 20 nm. The spring constant and the resonant frequency of the cantilever are about 2.8 N/m and 60 kHz, respectively. The NOS samples were prepared by thermal oxidation and low-pressure chemical vapor deposition (LPCVD) on *p*-type Si(001) wafers with layer thicknesses (nitride/oxide) of 30-35 Å/22-25 Å. Figure 9 shows a bit array of alternate positive (bright contrast) and negative (dark contrast) charges that were detected by the aforementioned EFM scheme. The average pitch between bits is about 36 nm, close to an areal density of 500 Gbits/in^2. These negative and positive charges were injected by using +8 V/−9 V pulses of 10-ms width. From the known trap density of electron or hole in Si_3N_4 film (~$1\text{-}5\times10^{18}$ cm^{-3}), we estimate that each charged bit contains less than 20 electrons or holes in nitride. Besides, defects on nitride/oxide interface could also act

as trapping sites which can also contribute to EFM signal. The property of total trap concentration fundamentally limits the bit areal density of charge storage because if there are fewer electrons/holes in a single bit, the EFM signal-to-noise ratio of a charged bit is reduced.

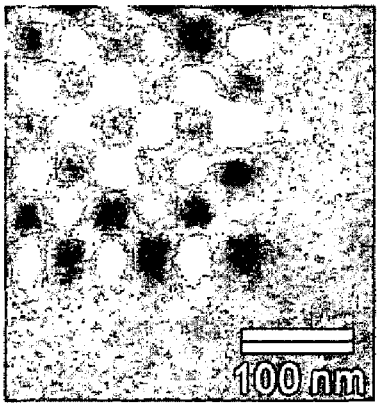

Figure 9 EFM readback signal of injected positive and negative charges.

4. Conclusions

For the MRM approach,. a CMOS-compatible fabrication process for the probe head has been demonstrated. The measured MR change after the structural release is lower than the common 2% value due the long routing leads through the mechanical springs. The tunneling magnetoresistive sensor can be a much better solution in terms of its high nominal resistance and large MR change. In regards to the tip position servo, we have demonstrated a stable displacement up to 60% of the gap against the well-known electrostatic pull-in by using a fixed linear controller. Based on our analysis, the traveled distance can be further extended to realize bit reading and writing, if a linear time-varying controller is available.

For the EFM approach, we have demonstrated the EFM reading and writing processes by using a conventional AFM instrument. Based on the similar dual-modulation scheme, chip-level cantilever sensing, actuation, and servo is expected to be demonstrated in the near future.

Acknowledgments

The authors wish to thank all the faculty and students in the CHIP center at Carnegie Mellon University for their contributions to this paper.

References:

[1] G. Binnig and H. Rohrer, US Patent 4,343,993 (1982).
[2] G. Binnig, US Patent 4,724, 318 (1988).
[3] A. Majumdar, *Annu. Rev. Mater. Sci.*, vol. 29, p. 505 (1999).
[4] L. Shi et al., *J. MEMS*, vol. 10, no. 3, p. 370 (2001).
[5] L. Kong, L. Zhuang, and S. Chou, *IEEE Trans. Magnetics*, vol. 33, no. 5, p. 3019 (1997).
[6] M. R. Koblischka et al., *Appl. Phys.* A, vol. 76, p. 879 (2003).
[7] P. Vettiger et al., *IEEE Trans. Nanotechnol.*, vol. 1, p. 39 (2002).
[8] L. R. Carley et al., *J. Appl. Phys.*, vol. 87, no. 9, p. 6680 (2000).
[9] L. Zhang, *IEEE Trans. Magnetics*, vol. 38, no. 5, p. 97 (2002).
[10] G. K. Fedder et al., *Sensors and Actuators A*, vol. 57, no. 2, p. 103 (1997).
[11] S. Min, J. A. Bain, and D. W. Greve, *J. Appl. Phys.*, vol. 91, no. 10, p. 6824 (2002).
[12] J. I. Seeger and B. E. Boser, *J. MEMS*, vol.12, p. 656-671, (2003).
[13] M. S.-C. Lu and G. K. Fedder, *J. MEMS*, to be published in August (2003).
[14] H. A. R. Wegener, A. J. Lincoln, H. C. Pao, M. R. O'Connel, and R. E. Oleksiak, *IEEE IEDM Tech. Dig.* , Washington, D.C. (1967).
[15] S. Iwamura, Y. Nishida, and K. Hashimoto, *IEEE Trans. Electron Devices*, 28, p. 854 (1981).
[16] R. C. Barrett and C. F. Quate, *J. Appl. Phys.*, 70, 2725 (1991).
[17] B. D. Terris and R. C. Barrett, *IEEE Trans. Electron Devices*, 42, 944 (1995).
[18] H. J. Mamin, B. D. Terris, L. S. Fan, S. Hoen, R. C. Barrett, D. Rugar, *IBM J. Res. Develp.*, 39, 681 (1995).
[19] I. Fujiwara, S. Kojima, and J. Seto, *Jap. J. Appl. Phys.* 35, 2764 (1996).
[20] W. M. D. Wright and D. G. Chetwynd, *Nanotechnology*, 9, 133 (1998).
[21] H. O. Jacobs and A. Stemmer, *Surf. Interface Anal.*, 27, 361 (1999).
[22] D. M. Schaadt, E. T. Yu, S. Sankar, and A. E. Berkowitz, *Appl. Phys. Lett.*, 74, p. 472 (1999).
[23] J. W. Hong, S. M. Shin, C. J. Kang, Y. Kuk, Z. G. Khim, and S. Park, *Appl. Phys. Lett.*, 75, p. 1760 (1999).
[24] G. H. Buh, H. J. Chung, Y. Kuk, *Appl. Phys. Lett.*, 79, p. 2010 (2001).
[25] S.-D. Tseng and S. Gwo, unpublished.
[26] S.-D. Tzeng, C.-L. Wu, Y.-C. You, T. T. Chen, S Gwo, and H. Tokumoto, *Appl. Phys. Lett.*, 81, p. 5042 (2002).
[27] F. S.-S. Chien, J.-W. Chang, S.-W. Lin, Y. C. Chou, T. T. Chen, S. Gwo, T.-S. Chao, and W.-F. Hsieh, *Appl. Phys. Lett.*, 76, p. 360 (2000).